Residual component analysis of hyperspectral images - Application to joint nonlinear unmixing and nonlinearity detection

Residual Component Analysis of Hyperspectral Images—Application to Joint Nonlinear Unmixing and Nonlinearity Detection

Yoann Altmann, *Member, IEEE*, Nicolas Dobigeon, *Senior Member, IEEE*, Steve McLaughlin, *Fellow, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*— This paper presents a nonlinear mixing model for joint hyperspectral image unmixing and nonlinearity detection. The proposed model assumes that the pixel reflectances are linear combinations of known pure spectral components corrupted by an additional nonlinear term, affecting the end members and contaminated by an additive Gaussian noise. A Markov random field is considered for nonlinearity detection based on the spatial structure of the nonlinear terms. The observed image is segmented into regions where nonlinear terms, if present, share similar statistical properties. A Bayesian algorithm is proposed to estimate the parameters involved in the model yielding a joint nonlinear unmixing and nonlinearity detection algorithm. The performance of the proposed strategy is first evaluated on synthetic data. Simulations conducted with real data show the accuracy of the proposed unmixing and nonlinearity detection strategy for the analysis of hyperspectral images.**

***Index Terms*— Hyperspectral imagery, nonlinear spectral unmixing, residual component analysis, nonlinearity detection.**

I. Introduction

SPECTRAL unmixing (SU) of hyperspectral images has attracted growing interest over the last few decades. It consists of distinguishing the materials and quantifying their proportions in each pixel of the observed image. This blind source separation problem has been widely studied for the applications where pixel reflectances are linear combinations of pure component spectra [1]–[5]. However, as explained in [6], [7], the linear mixing model (LMM) can be inappropriate for some hyperspectral images, such as those containing sand, trees or vegetation areas. Nonlinear mixing models (NLMMs) provide an interesting alternative to overcoming the inherent limitations of the LMM. They have been proposed in the hyperspectral image literature and can be divided into two main classes [8].

The first class of NLMMs consists of physical models based on the nature of the environment. These models include the bidirectional reflectance based model proposed in [9] for intimate mixtures associated with sand-like materials and the bilinear models recently studied in [10]–[13] to account for scattering effects mainly observed in vegetation and urban areas. The second class of NLMMs contains more flexible models allowing for different kinds of nonlinearities to be approximated. These flexible models are constructed from neural networks [14], [15], kernels [16], [17], or post-nonlinear transformations [18].

While the consideration of nonlinear effects can be relevant in specific areas, the LMM is often sufficient for approximating the actual mixing models in some image pixels, for instance in homogeneous regions. Thus, it makes sense to distinguish in any image, linearly mixed pixels which can be easily analyzed, from those nonlinearly mixed requiring deeper analysis. Nonlinearity detection in hyperspectral images has already been addressed in [19] to detect nonlinear areas in observed scenes using surrogate data. In previous work, a pixel-by-pixel nonlinearity detector based on a polynomial post-nonlinear mixing model (PPNMM) was proposed and provided interesting results [20]. The detector in [20] follows a PPNMM-based SU procedure and uses the statistical properties of the parameter estimator to subsequently derive an accurate test statistic. This paper proposes to simultaneously achieve the SU and nonlinearity detection. This problem has been recently addressed using sparse SU techniques based on bilinear [21] and post-nonlinear models [22], [23]. Conversely, we propose to use a model-selection approach for detecting nonlinearities with different statistical properties.

This paper presents a new supervised Bayesian algorithm for joint nonlinear SU and nonlinearity detection. This algorithm is supervised in the sense that the endmembers contained in the image are assumed to be known (chosen from a spectral library or extracted from the data by an endmember extraction algorithm (EEA)). This algorithm is based on a nonlinear mixing model inspired from residual component analysis (RCA) [24]. In the context of SU of hyperspectral images, the nonlinear effects are modeled by additive

Manuscript received June 27, 2013; revised December 2, 2013 and March 5, 2014; accepted March 6, 2014. Date of publication March 19, 2014; date of current version April 3, 2014. This work was supported in part by the Direction Générale de l'armement, French Ministry of Defence, in part by the Hypanema ANR under Project n_ANR-12-BS03-003, and in part by the Thematic Trimester on Image Processing of the CIMI Labex. The work of S. McLaughlin was supported by the Engineering and Physical Sciences Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. James E. Fowler.

Y. Altmann, N. Dobigeon, and J.-Y. Tourneret are with the University of Toulouse, Toulouse Cedex 7 31071, France (e-mail: yoann.altmann@enseeiht.fr; nicolas.dobigeon@enseeiht.fr; jean-yves.tourneret@enseeiht.fr).

S. McLaughlin is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH9 3JL, U.K. (e-mail: s.mclaughlin@hw.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2312616

perturbation terms characterized by Gaussian processes (GPs). This allows the nonlinear terms to be marginalized, yielding a flexible model depending only on the nonlinearity energies. The hyperspectral image to be analyzed is partitioned into homogeneous regions in which the nonlinearities share the same GP. This algorithm relies on an implicit image classification, modeled by labels whose spatial dependencies follow a Potts-Markov random field. Consideration of two classes (linear vs. nonlinear mixtures) would lead to binary detection maps. However, this paper allows for nonlinearly mixed regions to be also identified, based on the energy of the nonlinear effects. More precisely, the proposed algorithm can identify regions with different level of nonlinearity and characterized by different GPs. Most SU algorithms assume additive, independent and identically distributed (i.i.d.) noise sequences. However, based on previous work conducted on real hyperspectral images, non i.i.d. noise vectors are considered in this paper.

When the endmembers used to estimate the abundances are accurate, it makes sense to assume that the unknown abundances sum to one for each pixel (at least when assuming the LMM). However, this assumption can be relaxed. Two alternative algorithms are proposed in this paper (with and without the abundance sum-to-one constraint). However, only the fully constrained version is detailed for ease of reading. Details of the second version can be found in [25]. Moreover, it is also reasonable to assume that the nonlinearities only involve nonlinear transformations of the known endmembers. Modeling endmember estimation errors for supervised SU is an interesting problem that is however out of scope of this paper and a topic for future investigations.

In the Bayesian framework, appropriate prior distributions are chosen for the unknown parameters of the proposed RCA model, i.e., the mixing coefficients, the GP hyperparameters, the labels and the noise covariance matrix. The joint posterior distribution of these parameters is then derived. However, the classical Bayesian estimators cannot be easily computed from this joint posterior. To alleviate this problem, a Markov chain Monte Carlo (MCMC) method is used to generate samples according to the posterior of interest. Finally, the generated samples are used to compute Bayesian estimators as well as measures of uncertainties such as confidence intervals.

The remaining paper is organized as follows. Section II introduces the RCA model for hyperspectral image analysis. Section III presents the hierarchical Bayesian model associated with the proposed RCA model and its posterior distribution. The Metropolis-Within-Gibbs sampler used to sample from the posterior of interest is detailed in Section V. Some simulation results conducted on synthetic and real data are shown and discussed in Sections VI and VII. Conclusions are finally reported in Section VIII.

II. Problem Formulation

We consider a set of N observed pixel spectra $\mathbf{y}_n = [y_{n,1}, \ldots, y_{n,L}]^T, n \in \{1, \ldots, N\}$ where L is the number of spectral bands. Each of these spectra is defined as a linear combination of R known spectra $\mathbf{m}_r$, referred to as endmembers, contaminated by an additional spectrum $\boldsymbol{\phi}_n$ and additive noise

$$\begin{aligned}\mathbf{y}_n &= \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \boldsymbol{\phi}_n + \mathbf{e}_n \\ &= \mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n + \mathbf{e}_n, \quad n = 1, \ldots, N \end{aligned} \tag{1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene, $a_{r,n}$ is its corresponding proportion in the nth pixel and $\mathbf{e}_n$ is an additive independently and non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\boldsymbol{\Sigma}_0 = \text{diag}\left(\boldsymbol{\sigma}^2\right)$, denoted as $\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}_L, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and $\text{diag}\left(\boldsymbol{\sigma}^2\right)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. Moreover, the term $\boldsymbol{\phi}_n = [\phi_{1,n}, \ldots, \phi_{L,n}]^T$ in (1) is an unknown $L \times 1$ additive perturbation vector modeling nonlinear effects occurring in the nth pixel. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ have been used in the second row of Eq. (1). There are several motivations for considering the mixing model (1). First, 1) this model reduces to the classical linear mixing model (LMM) for $\boldsymbol{\phi}_n = \mathbf{0}_L$, 2) the model (1) is general enough to handle different of kinds of nonlinearities such as the bilinear model studied in [12] (referred to as Fan model (FM)), the generalized bilinear model (GBM) [13], and the polynomial post-nonlinear mixing model (PPNMM) studied for nonlinear spectral unmixing in [18] and nonlinearity detection in [20]. These models assume that the mixing model consists of a linear contribution of the endmembers, corrupted by at least one additive term characterizing the nonlinear effects. In the proposed model, all additive terms are gathered in the vector $\boldsymbol{\phi}_n$. Note that a similar model, called robust LMM, has been also introduced in [26].

Due to physical considerations, the abundance vectors $\boldsymbol{a}_n$ satisfy the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_{r,n} = 1, \quad a_{r,n} > 0, \forall r \in \{1, \ldots, R\}. \tag{2}$$

In this paper, the sum-to-one constraint is considered for the abundances. However, this constraint can be relaxed, as discussed in Section V-F. The problem addressed in this paper consists of the joint estimation of the abundance vectors and the detection of nonlinearly mixed pixels (characterized by $\boldsymbol{\phi}_n \neq \mathbf{0}_L$). The two next sections present the proposed Bayesian model for joint unmixing and nonlinearity detection.

III. Bayesian Linear Model

The unknown parameter vector associated with the proposed model (1) contains the abundances $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ (satisfying the constraints (2)), the nonlinear terms of each pixel $\left\{\boldsymbol{\phi}_n\right\}_{n=1,\ldots,N}$, and the noise variance vector $\boldsymbol{\sigma}^2$. This section summarizes the likelihood and the parameter priors associated with the parameters of the linear part of the model,

i.e., $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ and $\boldsymbol{\sigma}^2$. One of the main contributions of this paper is the characterization of the nonlinearities that will addressed later in Section IV.

A. Likelihood

Equation (1) shows that $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\phi}_n, \boldsymbol{\sigma}^2$ is distributed according to a Gaussian distribution with mean $\mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n$ and covariance matrix $\boldsymbol{\Sigma}_0$, denoted as $\mathbf{y}_n|\mathbf{M}, \boldsymbol{a}_n, \boldsymbol{\phi}_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}\left(\mathbf{M}\boldsymbol{a}_n + \boldsymbol{\phi}_n, \boldsymbol{\Sigma}_0\right)$. Assuming independence between the observed pixels, the joint likelihood of the observation matrix $\mathbf{Y}$ can be expressed as
$f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \boldsymbol{\Phi}, \boldsymbol{\sigma}^2)$

$$\propto |\boldsymbol{\Sigma}_0|^{-N/2} \text{etr}\left[-\frac{(\mathbf{Y}-\boldsymbol{X})^T\boldsymbol{\Sigma}_0^{-1}(\mathbf{Y}-\boldsymbol{X})}{2}\right] \tag{3}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]^T$ is an $L \times N$ nonlinearity matrix, $\propto$ means "proportional to", $\text{etr}(\cdot)$ denotes the exponential trace and $\boldsymbol{X} = \mathbf{M}\mathbf{A} + \boldsymbol{\Phi}$ is an $L \times N$ matrix.

B. Prior for the Abundance Matrix $\boldsymbol{A}$

Each abundance vector can be written as $\boldsymbol{a}_n = [\mathbf{c}_n^T, a_{R,n}]^T$ with $\mathbf{c}_n = [a_{1,n}, \ldots, a_{R-1,n}]^T$ and $a_{R,n} = 1 - \sum_{r=1}^{R-1} a_{r,n}$. The LMM constraints (2) impose that $\mathbf{c}_n$ belongs to the simplex

$$\mathcal{S} = \left\{ \mathbf{c} \,\middle|\, c_r > 0, \forall r \in 1, \ldots, R-1, \sum_{r=1}^{R-1} c_r < 1 \right\} \tag{4}$$

To reflect the lack of prior knowledge about the abundances, we propose to assign noninformative prior distributions for the N vectors $\mathbf{c}_n$. More precisely, the following uniform prior

$$f(\mathbf{c}_n) \propto \mathbf{1}_{\mathcal{S}}(\mathbf{c}_n), \quad n \in \{1, \ldots, N\} \tag{5}$$

is assigned for each vector $\mathbf{c}_n$, where $\mathbf{1}_{\mathcal{S}}(\cdot)$ is the indicator function defined on the simplex $\mathcal{S}$. Assuming prior independence between the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ leads to the following joint prior distribution

$$f(\mathbf{C}) = \prod_{n=1}^{N} f(\mathbf{c}_n) \tag{6}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ is an $(R-1) \times N$ matrix.

C. Prior for the Noise Variance Vector $\boldsymbol{\sigma}^2$

A noninformative Jeffreys' prior is chosen for the noise variance of each spectral band σ_ℓ^2

$$f(\sigma_\ell^2) \propto \frac{1}{\sigma_\ell^2}\mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right) \tag{7}$$

which reflects the absence of knowledge for this parameter (see [27] for motivation). Assuming prior independence between the noise variances, we obtain

$$f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f\left(\sigma_\ell^2\right). \tag{8}$$

IV. Modeling the Nonlinearities

We propose in this paper to exploit spatial correlations between the pixels of the hyperspectral image to be analyzed. It seems reasonable to assume that nonlinear effects occurring in a given pixel are related to the nonlinear effects present in neighboring pixels. Formally, the hyperspectral image is assumed to be partitioned into K classes denoted as $\mathcal{C}_0, \ldots, \mathcal{C}_{K-1}$. Let $\mathcal{I}_k \subset 1, \ldots, N$ denote the subset of pixel indexes belonging to the kth class ($k = 0, \ldots, K-1$). An $N \times 1$ label vector $\boldsymbol{z} = [z_1, \ldots, z_N]^T$ with $z_n \in \{0, \ldots, K-1\}$ is introduced to identify the class of each image pixel, i.e., $\mathbf{y}_n \in C_k \Leftrightarrow n \in \mathcal{I}_k \Leftrightarrow z_n = k$. In each class, the unknown nonlinearity vectors are assumed to share the same statistical properties, as will be shown in the sequel.

A. Prior Distribution for the Nonlinearity Matrix $\boldsymbol{\Phi}$

As mentioned above, the mixing model (1) reduces to the LMM for $\boldsymbol{\phi}_n = \mathbf{0}_L$. For nonlinearity detection, it makes sense to consider a pixel class (referred to as class $\mathcal{C}_0$) corresponding to linearly mixed pixels. The resulting prior distribution for $\boldsymbol{\phi}_n$ conditioned upon $z_n = 0$ is given by

$$f(\boldsymbol{\phi}_n|z_n = 0) = \prod_{\ell=1}^{L} \delta(\phi_{\ell,n}). \tag{9}$$

It can be seen that bilinear models and more generally polynomial models (i.e., model involving polynomials nonlinearities with respect to the endmembers) are particularly well adapted to model scattering effects, mainly observed in vegetation and urban areas. Consequently, it makes sense to assume that the nonlinearities $\boldsymbol{\phi}_n$ depend on the endmember matrix $\mathbf{M}$. Nonlinear effects can vary, depending on the relief of the scene, the underlying components involved in the mixtures and the observation conditions to name a few factors. This makes the choice of a single informative prior distribution challenging. From a classification point of view, it is interesting to identify regions or classes where similar nonlinearities occur. For these reasons, we propose to divide nonlinearly mixed pixels into $K-1$ classes and to assign different priors for the nonlinearity vectors belonging to the different classes. The nonlinearities (of nonlinearly mixed pixels) are assumed to be random. Assume $\mathbf{y}_n$ belongs to the kth class. The prior distribution of the corresponding nonlinear term $\boldsymbol{\phi}_n$ is given by the following GP ($k = 1, \ldots, K-1$)

$$\boldsymbol{\phi}_n|\mathbf{M}, z_n = k, s_k^2 \sim \mathcal{N}\left(\mathbf{0}_L, s_k^2\boldsymbol{K}_{\mathbf{M}}\right), \tag{10}$$

where $\boldsymbol{K}_{\mathbf{M}}$ is an $L \times L$ covariance matrix parameterized by the endmember matrix $\mathbf{M}$ and s_k^2 is a scaling hyperparameter that tunes the energy of the nonlinearities in the kth class. Note that all nonlinearity vectors within the same class share the same prior. The performance of the unmixing procedure depends on the choice of $\boldsymbol{K}_{\mathbf{M}}$, more precisely on the similarity measure associated with the covariance matrix. In this paper, we consider the symmetric second order polynomial kernel, which has received considerable interest in the machine learning community [28]. This kernel is defined as follows

$$[\boldsymbol{K}_{\mathbf{M}}]_{i,j} = \left(\mathbf{m}_{i,:}\mathbf{m}_{j,:}^T\right)^2, \quad i, j \in \{1, \ldots, L\}, \tag{11}$$

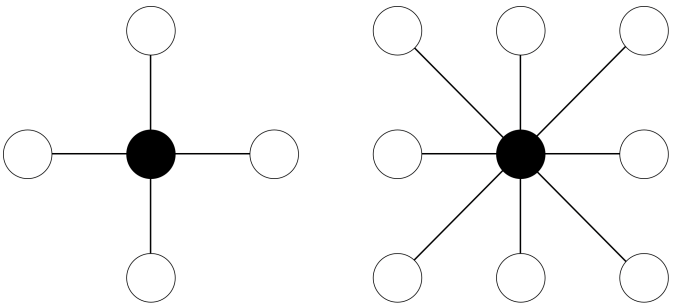

Fig. 1. 4-pixel (left) and 8-pixel (right) neighborhood structures. The considered pixel appear as a black circle whereas its neighbors are depicted in white.

where $\mathbf{m}_{i,:}$ is an $1 \times R$ vector that denotes the ith row of $\mathbf{M}$. Polynomial kernels are particularly well adapted to characterize multiple scattering effects (modeled by polynomial functions of the endmembers). Note that the parametrization of the matrix $\boldsymbol{K}_{\mathbf{M}}$ in (11) only involves bilinear and quadratic terms[1] with respect to the endmembers $\mathbf{m}_r, r = 1, \ldots, R$. More, precisely, the matrix $\boldsymbol{K}_{\mathbf{M}}$ can be rewritten as $\boldsymbol{K}_{\mathbf{M}} = \boldsymbol{Q}\boldsymbol{Q}^T$ where $\boldsymbol{Q} = [\mathbf{m}_1 \odot \mathbf{m}_1, \ldots, \mathbf{m}_R \odot \mathbf{m}_R, \sqrt{2}\mathbf{m}_1 \odot \mathbf{m}_2, \ldots, \sqrt{2}\mathbf{m}_{R-1} \odot \mathbf{m}_R]$ is an $L \times R(R+1)/2$ matrix and $\odot$ denotes the Hadamard (termwise) product. Note also that a polynomial kernel similar to (11) has been recently considered in [16] and that other kernels such as the Gaussian kernel could be investigated to model other nonlinearities as in [24]. As mentioned above, the endmembers of the scene are assumed to be known in this paper. Consequently, the proposed nonlinear model does not involved endmember estimation errors (i.e., missing or poorly estimated endmembers).

B. Prior Distribution for the Label Vector z

In the context of hyperspectral image analysis, the labels $z_1, \ldots, z_N$ indicate the pixel classes and take values in $\{0, \ldots, K-1\}$ where K is the number of classes and the set $\{z_n\}_{n=1,\ldots,N}$ forms a random field. To exploit the correlation between pixels, a Markov random field is introduced as a prior distribution for z_n given its neighbors $\boldsymbol{z}_{\mathcal{V}(n)}$, i.e., $f(z_n|\boldsymbol{z}_{\backslash n}) = f(z_n|\boldsymbol{z}_{\mathcal{V}(n)})$, where $\mathcal{V}(n)$ is the neighborhood of the nth pixel and $\boldsymbol{z}_{\backslash n} = \{z_{n'}\}_{n' \neq n}$. More precisely, this paper focuses on the Potts-Markov model since it is very appropriate for hyperspectral image segmentation [29]. Given a discrete random field $\boldsymbol{z}$ attached to an image with N pixels, the Hammersley-Clifford theorem yields

$$f(\boldsymbol{z}) = \frac{1}{G(\beta)} \exp\left[\beta \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n - z_{n'})\right]. \quad (12)$$

where $\beta > 0$ is the granularity coefficient, $G(\beta)$ is a normalizing (or partition) constant and $\delta(\cdot)$ is the Dirac delta function. Several neighborhood structures can be employed to define $\mathcal{V}(n)$. Fig. 1 shows two examples of neighborhood structures.

[1]Note: it can be shown that (10) and (11) can be obtained by defining $\boldsymbol{\phi}_n$ as a linear combination of terms $\mathbf{m}_i \odot \mathbf{m}_j$ (as in [13]) and marginalizing the corresponding coefficients using a Gaussian prior parameterized by s_k^2. Marginalizing these coefficients allows the number of unknown parameters to be significantly reduced, leading to the nonlinearities being characterized by a single parameter s_k^2.

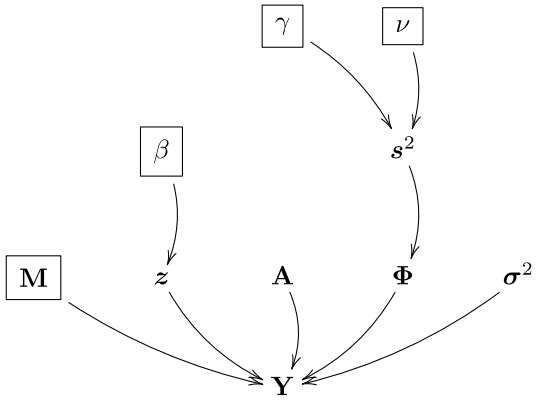

Fig. 2. DAG for the parameter and hyperparameter priors (the fixed parameters appear in boxes).

The four pixel structure (or 1-order neighborhood) will be considered in the rest of the paper.

The hyperparameter β tunes the degree of homogeneity of each region in the image. More precisely, small values of β yield an image with a large number of regions, whereas large values of β lead to fewer and larger homogeneous regions. In this paper, the granularity coefficient is assumed to be known. Note however that it could be also included within the Bayesian model and estimated using the strategy described in [30].

C. Hyperparameter Priors

The performance of the proposed Bayesian model for spectral unmixing mainly depends on the values of the hyperparameters $\{s_k^2\}_{k=1,\ldots,K}$. When the hyperparameters are difficult to adjust, it is the norm to include them in the unknown parameter vector, resulting in a hierarchical Bayesian model [18], [31]. This strategy requires the definition of prior distributions for the hyperparameters.

The following inverse-Gamma prior distribution

$$s_k^2|\gamma, \nu \sim \mathcal{IG}(\gamma, \nu), \quad \forall k \in \{1, \ldots, K\} \quad (13)$$

is assigned for the nonlinearity hyperparameters, where (γ, ν) are additional parameters that will be fixed to ensure a noninformative prior for s_k^2 ($(\gamma, \nu) = (1, 1/4)$ in all simulations presented in this paper). Assuming prior independence between the hyperparameters, we obtain

$$f(\boldsymbol{s}^2|\gamma, \nu) = \prod_{k=1}^{K-1} f(s_k^2|\gamma, \nu). \quad (14)$$

where $\boldsymbol{s}^2 = [s_1^2, \ldots, s_K^2]^T$.

V. Bayesian Inference Using a Metropolis-Within-Gibbs Sampler

A. Marginalized Joint Posterior Distribution

The resulting directed acyclic graph (DAG) associated with the proposed Bayesian model introduced in Sections III and IV is depicted in Fig. 2.

Assuming prior independence between $\mathbf{A}$, $(\boldsymbol{\Phi}, \boldsymbol{z})$ and $\boldsymbol{\sigma}^2$, the posterior distribution of $(\boldsymbol{\Phi}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\mathbf{C}, \boldsymbol{z}, \boldsymbol{\sigma}^2, \boldsymbol{s}^2)$

can be expressed as

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{Y}, \mathbf{M}) \propto f(\mathbf{Y}|\mathbf{M}, \boldsymbol{\theta}, \boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M}, z, s^2) f(\boldsymbol{\theta}),$$

where $f(\boldsymbol{\theta}) = f(\mathbf{C}) f(\boldsymbol{\sigma}^2) f(z) f(s^2)$. This distribution can be marginalized with respect to $\boldsymbol{\Phi}$ as follows

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M}) &\propto f(\boldsymbol{\theta}) \int f(\mathbf{Y}|\mathbf{M}, \boldsymbol{\theta}, \boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M}, z, s^2) \mathrm{d}\boldsymbol{\Phi} \\ &\propto f(\boldsymbol{\theta}) f(\mathbf{Y}|\mathbf{M}, \boldsymbol{\theta}) \end{aligned} \quad (15)$$

where

$$\begin{aligned} f(\mathbf{Y}|\mathbf{M}, \boldsymbol{\theta}) &= \int f(\mathbf{Y}|\mathbf{M}, \boldsymbol{\theta}, \boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\mathbf{M}, z, s^2) \mathrm{d}\boldsymbol{\Phi} \\ &\propto \prod_{k=0}^{K-1} \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right] \end{aligned} \quad (16)$$

with $\boldsymbol{\Sigma}_0 = \mathrm{diag}\left(\boldsymbol{\sigma}^2\right)$, $\boldsymbol{\Sigma}_k = s_k^2 \boldsymbol{K}_{\mathbf{M}} + \boldsymbol{\Sigma}_0$ $(k = 1, \ldots, K-1)$ and $\bar{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{M}\boldsymbol{a}_n$. The advantage of this marginalization is to avoid sampling the nonlinearity matrix $\boldsymbol{\Phi}$. Thus, the nonlinearities are fully characterized by the known endmember matrix, the class labels and the values of the hyperparameters in $\boldsymbol{s}^2 = [s_1^2, \ldots, s_K^2]^T$.

Unfortunately, it is difficult to obtain closed form expressions for standard Bayesian estimators associated with (15). In this paper, we propose to use efficient Markov Chain Monte Carlo (MCMC) methods to generate samples asymptotically distributed according to (15). The next part of this section presents the Gibbs sampler which is proposed to sample according to (15). The principle of the Gibbs sampler is to sample according to the conditional distributions of the posterior of interest [32, Chap. 10]. Due to the large number of parameters to be estimated, it makes sense to use a block Gibbs sampler to improve the convergence of the sampling procedure. More precisely, we propose to sample sequentially the N labels in $\boldsymbol{z}$, the abundance matrix $\mathbf{A}$, the noise variances $\boldsymbol{\sigma}^2$ and $\boldsymbol{s}^2$ using moves that are detailed in the next paragraphs.

B. Sampling the Labels

For the nth pixel ($n \in \{1, \ldots, N\}$), the label z_n is a discrete random variable whose conditional distribution is fully characterized by the probabilities

$$\begin{aligned} P(z_n = k|\mathbf{y}_n, \mathbf{M}, \boldsymbol{\theta}_{\backslash z_n}) &\propto f(\mathbf{y}_n|\mathbf{M}, \boldsymbol{s}^2, z_n = k, \boldsymbol{a}_n) \\ &\quad \times f(z_n|\boldsymbol{z}_{\backslash n}), \end{aligned}$$

where $\boldsymbol{\theta}_{\backslash z_n}$ denotes $\boldsymbol{\theta}$ without z_n, $k = 0, \ldots, K-1$ (for K classes). These posterior probabilities can be expressed as

$$\begin{aligned} P(z_n = k|\mathbf{y}_n, \mathbf{M}, \boldsymbol{\theta}_{\backslash z_n}) &\propto \exp\left[\beta \sum_{p=1}^{N} \sum_{p' \in \mathcal{V}(p)} \delta(z_p - z_{p'})\right] \\ &\quad \times \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right]. \end{aligned} \quad (17)$$

Consequently, sampling z_n from its conditional distribution can be achieved by drawing a discrete value in the finite set $\{0, \ldots, K-1\}$ with the probabilities defined in (17).

C. Sampling the Abundance Matrix $\boldsymbol{A}$

Sampling from $f(\mathbf{C}|\mathbf{Y}, \mathbf{M}, \boldsymbol{\theta}_{\backslash \mathbf{C}})$ seems difficult due to the complexity of this distribution. However, it can be shown that

$$f(\mathbf{C}|\mathbf{Y}, \mathbf{M}, z, \boldsymbol{\sigma}^2, s^2) = \prod_{n=1}^{N} f(\mathbf{c}_n|\mathbf{y}_n, \mathbf{M}, z_n, \boldsymbol{\sigma}^2, s^2), \quad (18)$$

i.e., the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ are a posteriori independent and can be sampled independently in a parallel manner. Straightforward computations lead to

$$\mathbf{c}_n|\mathbf{y}_n, \mathbf{M}, z_n = k, \boldsymbol{\sigma}^2, s^2 \sim \mathcal{N}_{\mathcal{S}}(\bar{\boldsymbol{c}}_n, \boldsymbol{\Psi}_n) \quad (19)$$

where

$$\begin{aligned} \boldsymbol{\Psi}_n &= \left(\widetilde{\mathbf{M}}^T \boldsymbol{\Sigma}_k^{-1} \widetilde{\mathbf{M}}\right)^{-1} \\ \bar{\boldsymbol{c}}_n &= \boldsymbol{\Psi}_n \widetilde{\mathbf{M}}^T \boldsymbol{\Sigma}_k^{-1} \tilde{\mathbf{y}}_n \\ \widetilde{\mathbf{M}} &= [\mathbf{m}_1 - \mathbf{m}_R, \ldots, \mathbf{m}_{R-1} - \mathbf{m}_R] \end{aligned} \quad (20)$$

and $\tilde{\mathbf{y}}_n = \mathbf{y}_n - \mathbf{m}_R$. Moreover, $\mathcal{N}_{\mathcal{S}}(\bar{\boldsymbol{c}}_n, \boldsymbol{\Psi}_n)$ denotes the truncated multivariate Gaussian distribution defined on the simplex $\mathcal{S}$ with hidden mean $\bar{\boldsymbol{c}}_n$ and hidden covariance matrix $\boldsymbol{\Psi}_n$. Sampling from (19) can be achieved efficiently using the method recently proposed in [33].

D. Sampling the Noise Variance σ^2

It can be shown from (15) that

$$f(\boldsymbol{\sigma}^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, z, s^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, z, s^2), \quad (21)$$

where

$$\begin{aligned} &f(\sigma_\ell^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, z, s^2) \\ &\propto \frac{1}{\sigma_\ell^2} \prod_{k=0}^{K-1} \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right] \mathbf{1}_{\mathbb{R}^+}\left(\sigma_\ell^2\right) \end{aligned} \quad (22)$$

Sampling from (22) is not straightforward. In this case, an accept/reject procedure can be used to update σ_ℓ^2, leading to a hybrid Metropolis-within-Gibbs sampler. In this paper, we introduce the standard change of variable $\delta_\ell = \log(\sigma_\ell^2)$, $\delta_\ell \in \mathbb{R}$. A Gaussian random walk for δ_ℓ is used to update the variance σ_ℓ^2. Note that the noise variances are a posteriori independent. Thus they can be updated in a parallel manner. The variances of the L parallel Gaussian random walk procedures have been adjusted during the burn-in period of the sampler to obtain an acceptance rate close to 0.5, as recommended in [34, p. 8].

E. Sampling the Vector $\boldsymbol{s}^2$

It can be shown from (15) that

$$f(\boldsymbol{s}^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, z, \boldsymbol{\sigma}^2, \gamma, \nu) = \prod_{k=1}^{K-1} f(s_k^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \boldsymbol{\sigma}^2, \gamma, \nu),$$

where

$$\begin{aligned} &f(s_k^2|\mathbf{Y}, \mathbf{M}, \mathbf{A}, \boldsymbol{\sigma}^2, \gamma, \nu) \\ &\propto f(s_k^2|\gamma, \nu) \prod_{n \in \mathcal{I}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \bar{\mathbf{y}}_n^T \boldsymbol{\Sigma}_k^{-1} \bar{\mathbf{y}}_n\right]. \end{aligned} \quad (23)$$

Algorithm 1 Gibbs Sampling Algorithm

1: Fixed input parameters: $\mathbf{M}, \beta, K, \gamma, \nu$
2: Initialization $(t = 0)$
  • Set $\mathbf{A}^{(0)}, \boldsymbol{z}^{(0)}, \boldsymbol{\sigma}^{2(0)}, \boldsymbol{s}^{2(0)}$
3: Iterations $(t \geq 1)$
4: **for** $n = 1 : N$ **do**
5: Sample z_n from the probabilities in (17)
6: **end for**
7: Sample $\mathbf{A}^{(t)}$ from the pdf in (18)
8: Sample $\boldsymbol{\sigma}^{2(t)}$ from the pdfs in (22) and Gaussian random walks
9: Sample $\mathbf{s}^{2(t)}$ from the pdfs in (23) and Gaussian random walks
10: Set $t = t + 1$.

Due to the complexity of the conditional distribution (23), Gaussian random walk procedures are used in the log-space to update the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ in a parallel manner (similarly to the noise variance updates). Again, the proposal variances are adjusted during the burn-in period of the sampler. The resulting Metropolis-within-Gibbs sampler used to sample according to the posterior (15) is summarized in Algo. 1.

After generating N_{MC} samples using the procedures detailed above and removing N_{bi} iterations associated with the burn-in period of the sampler (N_{bi} has been set from preliminary runs), the marginal maximum a posteriori (MAP) estimator of the label vector, denoted as $\hat{\boldsymbol{z}}_{\text{MAP}}$, can be computed. The label vector estimator is then used to compute the minimum mean square error (MMSE) of $\mathbf{A}$ conditioned upon $\boldsymbol{z} = \hat{\boldsymbol{z}}_{\text{MAP}}$. Finally, the noise variances and the hyperparameters $\{s_k^2\}_{k=1,\ldots,K-1}$ are estimated using the empirical averages of the generated samples (MMSE estimates).

F. Relaxation of the Abundance Constraints

In this paper, the abundances are assumed to sum to one. This choice has been motivated by the fact that this constraint has been widely used for linear and nonlinear mixing models [12], [13], [16], [18]. However, the sum-to-one constraint can be removed when considering nonlinear mixtures, as proposed in [35]. In a Bayesian framework, relaxing the abundance sum-to-one constraint can be achieved by assigning a different prior for the abundances. An extension of the proposed algorithm has been investigated to relax the abundance sum-to-one constraints. For brevity, the Bayesian model and corresponding sampler have been omitted in this paper and have been reported in [25].

VI. Simulations for Synthetic Data

This section studies the performance of the proposed algorithm for synthetic hyperspectral images.

A. First Scenario: RCA vs. Linear Unmixing

The performance of the proposed joint nonlinear SU and nonlinearity detection algorithm is first evaluated by unmixing a synthetic image of 60×60 pixels generated according to the model (1). The $R = 3$ endmembers contained in these images (i.e., green grass, olive green paint and galvanized steel metal) have $L = 207$ different spectral bands and have been extracted from the spectral libraries provided with the ENVI software [36]. The number of classes has been set to $K = 4$, i.e, $K - 1 = 3$ classes of nonlinearly mixed pixels. The hyperparameters $\{s_k^2\}_{k=1,\ldots,3}$ have been fixed as shown in Table II, which represents three possible levels of nonlinearity. For each class, the nonlinear terms have been generated according to (10). The label map generated with $\beta = 1.6$ is shown in Fig. 3 (left). The abundance vectors $\boldsymbol{a}_n, n = 1, \ldots, 3600$ have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. The noise variance (depicted in Fig. 4 as a function of the spectral bands) have been arbitrarily fixed using $\sigma_\ell^2 = 10^{-4}\left[2 - \sin\left(\pi \frac{\ell}{L-1}\right)\right]$. to model a non-i.i.d. (colored) noise. The joint nonlinear SU and nonlinearity detection algorithm, denoted as "RCA-SU", has been applied to this data set with $N_{\text{MC}} = 4000$ and $N_{\text{bi}} = 2500$. Fig. 3 (right) shows that the estimated label map (marginal MAP estimates) is in agreement with the actual label map. Moreover, the confusion matrix depicted in Table I illustrate the performance of the RCA-SU in term of pixel classification. Table II shows that the RCA-SU provides accurate hyperparameter estimates and thus can be used to obtain information about the importance of nonlinearities

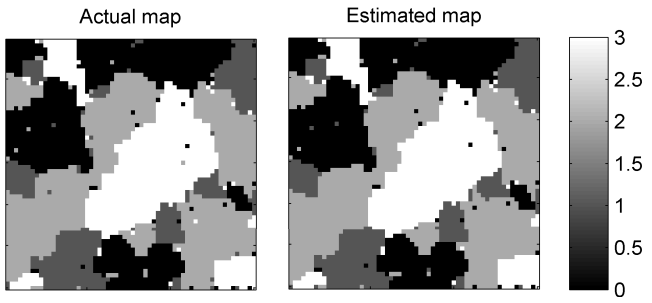

Fig. 3. Actual (left) and estimated (right) classification maps of the synthetic image associated with the first scenario.

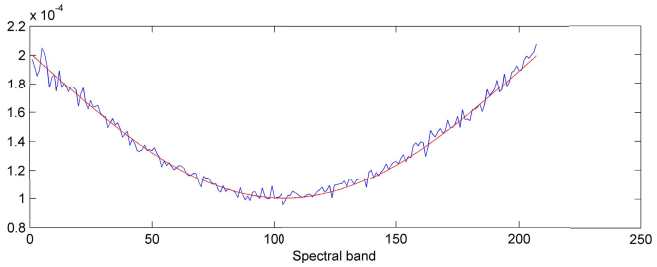

Fig. 4. Actual noise variances (red) and variances estimated by the RCA-SU algorithm (blue) for the synthetic image associated with the first scenario.

TABLE I
First Scenario: Confusion Matrix ($N = 3600$ Pixels)

		Estimated classes			
		$\mathcal{C}_0$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$
Actual classes	$\mathcal{C}_0$	984	1	0	0
	$\mathcal{C}_1$	0	520	9	0
	$\mathcal{C}_2$	1	7	1249	1
	$\mathcal{C}_3$	0	0	6	822

TABLE II
FIRST SCENARIO: HYPERPARAMETER ESTIMATION

	s_1^2	s_2^2	s_3^2
Actual value	0.01	0.1	1
Estimation error	6.69%	0.52%	2.00%

TABLE III
RNMSEs ($\times 10^{-2}$): SYNTHETIC IMAGES

	Class #0	Class #1	Class #2	Class #3
FCLS	0.39	15.33	29.63	42.53
RCA-SU	**0.38**	**2.77**	**3.96**	**4.50**

in the different regions. Note that the estimation error is computed using $|s_k^2 - \hat{s}_k^2|/s_k^2$, where s_k^2 and $\hat{s}_k^2$ are the actual and estimated dispersion parameters for the kth class. The estimated noise variances, depicted in Fig. 4 are also in good agreement with the actual values of the variances.

The quality of abundance estimation can be evaluated by comparing the estimated and actual abundance vectors using the root normalized mean square error (RNMSE) defined in each class by

$$\text{RNMSE}_k = \sqrt{\frac{1}{N_k R} \sum_{n \in \mathcal{I}_k} \left\| \hat{\boldsymbol{a}}_n - \boldsymbol{a}_n \right\|^2} \tag{24}$$

with $N_k = \text{card}(\mathcal{I}_k)$ and where $\boldsymbol{a}_n$ and $\hat{\boldsymbol{a}}_n$ are the actual and estimated abundance vectors for the nth pixel of the image. For this scenario, the proposed algorithm is compared with the classical FCLS algorithm [2] assuming the LMM. Comparisons to nonlinear SU methods will be addressed in the next paragraph (scenario 2). Table III shows the RNMSEs obtained with the proposed and the FLCS algorithms for this first data set. These results show that the two algorithms provide similar abundance estimates for the first class, corresponding to linearly mixed pixels. For the three nonlinear classes, the estimation performance is reduced. However, the proposed algorithm provides better results than the FCLS algorithm that does not handle nonlinear effects.

B. Second Scenario: RCA vs. Nonlinear Unmixing

1) Data Set: The performance of the proposed joint nonlinear SU and nonlinearity detection algorithm is then evaluated on a second synthetic image of 60×60 pixels containing the $R = 3$ spectral components presented in the previous section. In this scenario, the image consists of pixels generated according to four different mixing models associated with four classes ($K = 4$). The label map generated using $\beta = 1.6$ is shown in Fig. 5(a). The class $\mathcal{C}_0$ is associated with the LMM. The pixels of class $\mathcal{C}_1$ have been generated according to the generalized bilinear mixing model (GBM) [13]

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j} a_{i,n} a_{j,n} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n \tag{25}$$

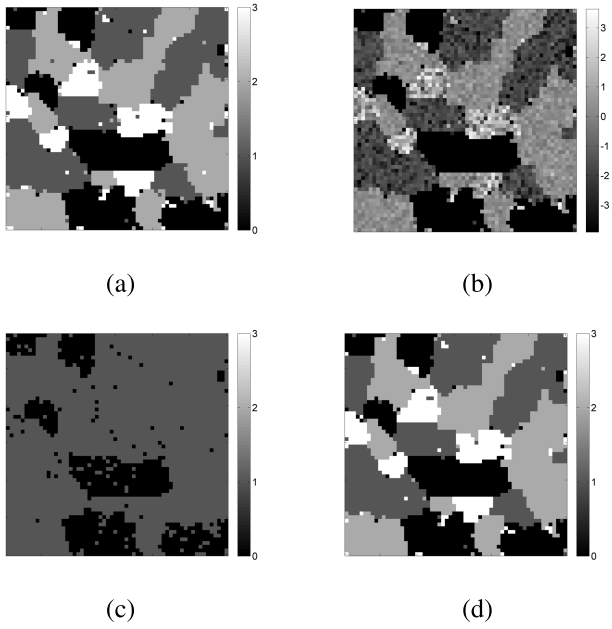

Fig. 5. Nonlinearity detection for the scenario #2. (a) Actual label map. (b) $\log\left(\|\boldsymbol{\phi}_n\|^2\right)$. (c) Detection map (PPNMM). (d) Detection map (RCA-SU).

where $n \in \mathcal{I}_1$ and the nonlinearity parameters $\{\gamma_{i,j}\}$ have been uniformly drawn in [0.5, 1]. The class $\mathcal{C}_2$ is composed of pixels generated according to the PPNMM [18] as follows

$$\mathbf{y}_n = \sum_{r=1}^{R} a_{r,n}\mathbf{m}_r + b\left(\sum_{r=1}^{R} a_{r,n}\mathbf{m}_r\right) \odot \left(\sum_{r=1}^{R} a_{r,n}\mathbf{m}_r\right) + \mathbf{e}_n \tag{26}$$

where $n \in \mathcal{I}_2$ and $b = 0.5$ for all pixels in class $\mathcal{C}_2$. Finally, the class $\mathcal{C}_3$ has been generated according to (1) with $s^2 = 0.1$. For the four classes, the abundance vectors have been randomly generated according to a uniform distribution over the admissible set defined by the positivity and sum-to-one constraints. All pixels have been corrupted by an additive i.i.d Gaussian noise of variance $\sigma^2 = 10^{-4}$, corresponding to an average signal-to-noise ratio SNR $\simeq$ 30dB. The noise is assumed to be i.i.d. for a fair comparison with SU algorithms assuming i.i.d. Gaussian noise. Fig. 5(b) shows the log-energy of the nonlinearity parameters for each pixel of the image, i.e., $\log\left(\|\boldsymbol{\phi}_n\|^2\right)$ for $n = 1, \ldots, 3600$. This figure shows that each class corresponds to a different level of nonlinearity.

2) Unmixing: Different estimation procedures have been considered for the four different mixing models:

- The FCLS algorithm [2] which is known to have good performance for linear mixtures (with the regularization parameter δ set to $\delta = 10^5$).
- The GBM-based approach [37] which is particularly adapted for bilinear nonlinearities. The optimization algorithm is stopped when the norm of the difference between consecutive parameter estimates is smaller than 10^{-6}.
- The gradient-based approach of [18] which is based on a PPNMM and has shown nice properties for various

TABLE IV
ABUNDANCE RNMSEs ($\times 10^{-2}$): SCENARIO #2

Unmixing algo.	Class #0 (LMM)	Class #1 (GBM)	Class #2 (PPNMM)	Class #3 (RCA)
FCLS	**0.35**	9.02	20.45	29.23
GBM	**0.36**	3.29	15.51	28.16
PPNMM	0.66	**1.38**	**0.47**	24.03
K-HYPE	3.36	3.26	3.13	**3.57**
RCA-SU	0.44	**1.55**	**2.13**	**3.52**

TABLE V
REs ($\times 10^{-2}$): SCENARIO #2

Unmixing algo.	Class #0 (LMM)	Class #1 (GBM)	Class #2 (PPNMM)	Class #3 (RCA)
FCLS	**0.99**	2.17	1.33	**3.10**
GBM	1.00	1.15	4.38	11.94
PPNMM	1.00	1.01	**0.99**	3.79
K-HYPE	**0.98**	**0.98**	**0.98**	**0.98**
RCA-SU	1.00	**0.99**	**0.98**	**0.98**

nonlinear models. This iterative algorithm is stopped when the difference of consecutive cost function values is smaller than 10^{-12}.

- The proposed RCA-SU algorithm which has been designed for the model in (1). It has been applied to this data set with $N_{\text{MC}} = 4000$, $N_{\text{bi}} = 2500$, $K = 4$ and $\beta = 1.6$.
- Finally, we consider the K-Hype method [16] to compare our algorithm with state-of-the art kernel based unmixing methods. The kernel used in this paper is the polynomial, second order symmetric kernel whose Gram matrix is defined by (11). This kernel provides better performance on this data set than the kernels studied in [16] (namely the Gaussian and the polynomial, second order asymmetric kernels). All hyperparameters of the K-Hype algorithm have been optimized using preliminary runs.

Table IV compares the RNMSEs obtained with the SU algorithms for each class of the second scenario. These results shows that the proposed algorithm provides abundance estimates similar to those obtained with the LMM-based algorithm (FCLS) for linearly mixed pixels. Moreover, the RCA-SU also provides accurate estimates for the three mixing models considered, which illustrates the robustness of the RCA-based model regarding model mis-specification.

The unmixing quality is also evaluated by the reconstruction error (RE) defined as $\text{RE}_k = \sqrt{\sum_{n \in \mathcal{I}_k} \left\| \hat{\mathbf{y}}_n - \mathbf{y}_n \right\|^2 / (N_k L)}$, where $\mathbf{y}_n$ is the nth observation vector and $\hat{\mathbf{y}}_n$ its estimate. Table V compares the REs obtained for the different classes. This table shows the accuracy of the proposed model for fitting the observations. The REs obtained with the RCA-SU are similar for the four pixel classes. Moreover, the performance in terms of RE of the proposed algorithm are similar to the performance of the K-Hype algorithm. Table VI compares the processing time of the different unmixing algorithms considered to process the synthetic data of the second scenario. This table shows that the proposed algorithm requires a higher computational cost when compared to the other algorithms, mainly due to the sampling procedure. However, it is important to note that since the proposed hybrid Gibbs sampler is highly parallelizable, (i.e., the N abundance vectors are a posteriori independent and the label vector can be efficiently updated using two sequential updates for a 4-pixel neighborhood), it does not suffer from potential computational burden induced by processing the image pixels sequentially.

TABLE VI
PROCESSING TIME (IN S): SCENARIO #2

FCLS	GBM	PPNMM	K-HYPE	RCA-SU
1	773	215	9	1244

From a reconstruction point of view, the K-Hype and RCA-SU algorithms provides similar results. However, the proposed algorithm also provides nonlinearity detection maps. The PPNMM and RCA-SU algorithms perform similarly in term of abundance estimation and allow both nonlinearities to be detected in each pixel. However, the nonlinearities can be analyzed more deeply using the RCA-SU, as will be shown in the next part.

3) Nonlinearity Detection: The performance of the proposed algorithm for nonlinearity detection is compared to the detector studied in [20], which is coupled with the PPNMM-based SU procedure mentioned above. The probability of false alarm of the PPNMM-based detection has been set to PFA = 0.05. Fig. 5(c) and (d) show the detection maps obtained with the two detectors. Both detectors are able to locate the nonlinearly mixed regions. However, the RCA-SU provides more homogeneous regions, due to the consideration of spatial structure through the MRF. Moreover, the proposed algorithm provides information about the different levels of nonlinearity in the image thanks to the estimation of the hyperparameters s_k^2 associated with the different classes. In this simulation, we obtain $[\hat{s}_1^2, \hat{s}_2^2, \hat{s}_3^2] = [0.2, 1.3, 10] \times 10^{-2}$, showing that nonlinearities of class $\mathcal{C}_1$ are less severe than those of class $\mathcal{C}_2$ and that are themselves weaker than those of class $\mathcal{C}_3$. The next section studies the performance of the proposed algorithm for a real hyperspectral image.

VII. SIMULATIONS FOR A REAL HYPERSPECTRAL IMAGE

A. Data Set

The real image considered in this section was acquired in 2010 by the Hyspex hyperspectral scanner over Villelongue, France (00°03'W and 42°57'N). $L = 160$ spectral bands were recorded from the visible to near infrared with a spatial resolution of 0.5m. This dataset has already been studied in [17] and [38] and is mainly composed of forested and urban areas. More details about the data acquisition and preprocessing steps are available in [38]. A sub-image of size 180×250 pixels is chosen here to evaluate the proposed unmixing procedure and is depicted in Fig. 6. The scene is composed mainly of a path and different vegetation species, resulting in $R = 5$ endmembers. The spectral signatures of

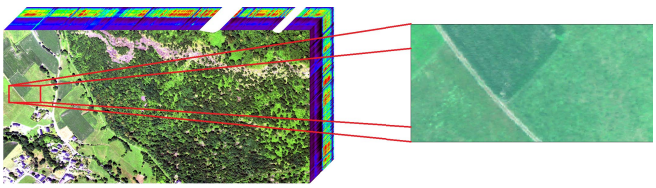

Fig. 6. Real hyperspectral Madonna data acquired by the Hyspex hyperspectral scanner over Villelongue, France (left) and sub-image of interest (right).

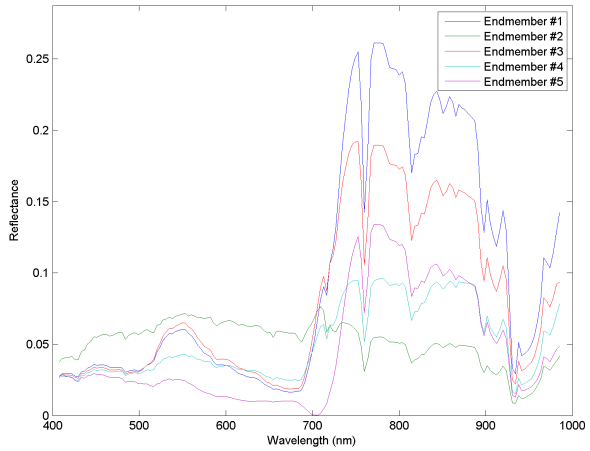

Fig. 7. The $R = 5$ endmembers estimated by the LMM-based algorithm [31] for the real sub-image.

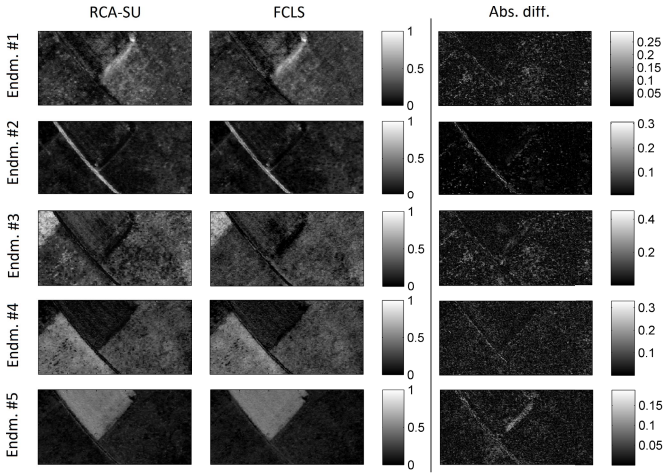

Fig. 8. The $R = 5$ abundance maps estimated by the RCA-SU (left) and FCLS (middle) algorithms for the Madonna real image (white pixels correspond to large abundances, contrary to black pixels). Right: Maps of absolute differences between the FCLS and RCA-SU abundance estimates.

the components have been extracted from the data using the LMM-based algorithm studied in [31] and are depicted in Fig. 7.

B. Spectral Unmixing

The proposed algorithm has been applied to this data set with $N_{\mathrm{MC}} = 4000$ and $N_{\mathrm{bi}} = 2500$. The number of classes has been set to $K = 5$ (one linear class and four nonlinear classes). The granularity parameter of the label prior (12) has been fixed to $\beta = 1.6$. Fig. 8 shows the abundance maps estimated by the FCLS algorithm and the proposed method. The abundance maps estimated by the RCA-SU algorithm

TABLE VII
RECONSTRUCTION ERRORS: REAL IMAGE

Unmixing algo.	RE ($\times 10^{-2}$)
FCLS	0.22
GBM	0.22
PPNMM	0.22
K-HYPE	0.22
RCA-SU	0.22

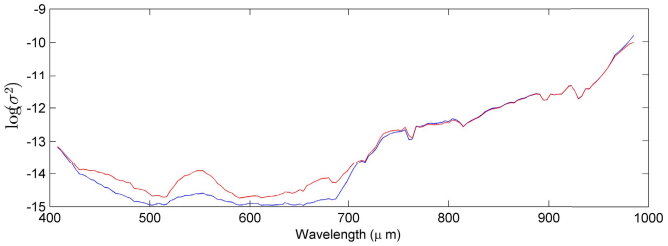

Fig. 9. Noise variances estimated by the RCA-SU (red) and the Hysime algorithm (blue) for the real Madonna image.

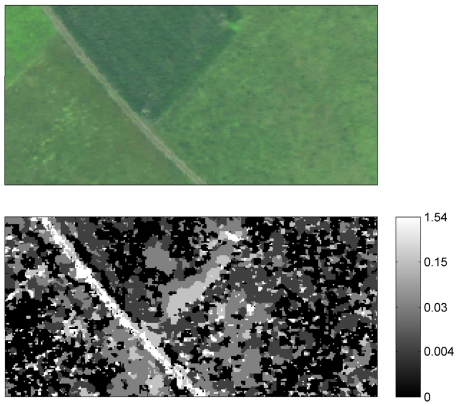

Fig. 10. Top: true color image of the scene of interest. Bottom: nonlinearity detection map obtained with the RCA-SU detector for the Madonna image ($K = 5$).

are in good agreement with those estimated by FCLS for most of the pixels but can differ locally. Table VII shows that the state-of-the-art and the proposed algorithm provide similar reconstruction errors. Fig. 9 compares the noise variance estimated by the RCA-SU for the real image with the noise variance estimated by the HySime algorithm [39]. This figure shows that the two algorithms provide similar noise variance estimates. These results motivate the consideration of non i.i.d. noise for hyperspectral image analysis since the noise variances increase for the highest wavelengths. The simulations conducted on this real dataset show the accuracy of the proposed RCA-SU in terms of abundance estimation and reconstruction error, especially for applications where the noise variances vary depending on the wavelength. Moreover, it also provides information about the nonlinearities of the scene.

C. Nonlinearity Detection

Fig. 10 (bottom) shows the detection map (map of z_n for $n = 1, \ldots, N$) provided by the proposed RCA-SU detector for

the real image considered. Due to the consideration of spatial structures, the proposed detector provides homogeneous regions. Similar structures can be identified in this detection map and the true color image of the scene [Fig. 10 (top)]. Moreover, the RCA-SU can identify four levels of nonlinearity, corresponding to $[\hat{s}_1^2, \hat{s}_2^2, \hat{s}_3^2, \hat{s}_4^2] = [0.004; 0.03; 0.15; 1.54]$. The estimated class $\mathcal{C}_4$ (white pixels) associated with the highest level of nonlinearity is mainly located on the path crossing the image. A second region of average nonlinearity level associated with the class $\mathcal{C}_3$ (light grey pixels) is mainly located in the pixels containing the first endmember. Finally, weak nonlinearities (classes $\mathcal{C}_2$ and $\mathcal{C}_1$) and linear mixtures (class $\mathcal{C}_0$) are located in homogeneous regions of the image. Additional simulation results conducted with different numbers of classes can be found in [25].

VIII. Conclusion

We have proposed a new hierarchical Bayesian algorithm for joint linear/nonlinear spectral unmixing of hyperspectral images and nonlinearity detection. This algorithm assumed that each pixel of the image is a linear or nonlinear mixture of endmembers contaminated by additive Gaussian noise. The nonlinear mixtures are decomposed into a linear combination of the endmembers and an additive term representing the nonlinear effects. A Markov random field was introduced to promote spatial structures in the image. The image was decomposed into regions or classes where the nonlinearities share the same statistical properties, each class being associated with a level of nonlinearity. Nonlinearities within a same class were modeled using a Gaussian process parametrized by the endmembers and the nonlinearity level. Note finally that the physical constraints for the abundances were included in the Bayesian framework through appropriate prior distributions. Due to the complexity of the resulting joint posterior distribution, a Markov chain Monte Carlo method was investigated to compute Bayesian estimators of the unknown model parameters.

Simulations conducted on synthetic data illustrated the performance of the proposed algorithm for linear and nonlinear spectral unmixing. An important advantage of the proposed algorithm is its robustness regarding the actual underlying mixing model. Another interesting property resulting from the nonlinear mixing model considered is the possibility of detecting several kinds of linearly and nonlinearly mixed pixels. This detection can be used to identify the image regions affected by nonlinearities in order to characterize the nonlinear effects more deeply. Finally, simulations conducted with real data showed the accuracy of the proposed unmixing and nonlinearity detection strategy for the analysis of real hyperspectral images.

The endmembers contained in the hyperspectral image were assumed to be known in this work. Of course, the performance of the algorithm relies on this endmember knowledge. We think that estimating the pure component spectra present in the image, jointly with the abundance estimation and the nonlinearity detection is an important issue that should be considered in future work. The number of classes and the granularity of the scene were assumed to be known in this study. Estimating these parameters is clearly a challenging issue that is under investigation. Finally, an extended algorithm has been proposed to estimate the abundances without abundance sum-to-one constraint, as often considered for images with significant shadowing effects. Modeling shadow in hyperspectral images is also a interesting prospect.

References

[1] M. Craig, "Minimum volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[2] D. C. Heinz and C.-I. Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 2001.

[3] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.

[4] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1008–1021, Apr. 2007.

[5] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[6] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[7] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[8] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[9] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, 1981.

[10] B. Somers *et al.*, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1183–1193, 2009.

[11] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," *Proc. SPIE*, vol. 7477, p. 74770I, Sep. 2009.

[12] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Remote Sens. Environ.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

[13] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[14] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 39, no. 8, pp. 2314–2318, Aug. 2001.

[15] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Proc. IEEE IGARSS*, Jul. 2011, pp. 1151–1154.

[16] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[17] Y. Altmann, N. Dobigeon, S. McLaughlin, and J. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[18] Y. Altmann, A. Halimi, N. Dobigeon, and J. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, Jun. 2012.

[19] T. Han and D. Goodenough, "Investigation of nonlinearity in hyperspectral imagery using surrogate data methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2840–2847, Oct. 2008.

[20] Y. Altmann, N. Dobigeon, and J. Tourneret, "Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1267–1276, Apr. 2013.
[21] Q. Qu, N. Nasrabadi, and T. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.
[22] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Bayesian algorithm for unsupervised unmixing of hyperspectral images using a post-nonlinear model," in *Proc. EUSIPCO*, Marrakech, Morrocco, Sep. 2013, pp. 1–5.
[23] Y. Altmann, N. Dobigeon, and J. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a hamiltonian monte carlo algorithm," *IEEE Trans. Image Process.*, accepted for publication.
[24] A. A. Kalaitzis and N. D. Lawrence, "Residual components analysis," in *Proc. ICML*, 2012, pp. 1–3.
[25] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images—Application to joint nonlinear unmixing and nonlinearity detection," IRIT-SC group, Univ. Toulouse, France, Tech. Rep., Sep. 2013.
[26] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS WHISPERS*, Jun. 2013, pp. 1–4.
[27] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York, NY, USA: Wiley, 1994.
[28] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
[29] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.
[30] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity coefficient of a Potts–Markov random field within an MCMC algorithm," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2385–2397, Jun. 2013.
[31] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 2657–2669, Nov. 2009.
[32] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer-Verlag, 2004.
[33] A. Pakman and L. Paninski, "Exact hamiltonian Monte Carlo for truncated multivariate Gaussians," *ArXiv e-prints*, Aug. 2012.
[34] C. P. Robert and D. Cellier, "Convergence control of MCMC algorithms," in *Discretization MCMC Convergence Assessment*, C. P. Robert, Ed. New York, NY, USA: Springer-Verlag, 1998, pp. 27–46.
[35] I. Meganem, P. Deliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear-quadratic mixing model for reflectances in urban environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 544–558, Jan. 2013.
[36] *ENVI User's Guide Version 4.0*, RSI Research Systems Inc., Boulder, CO, USA, Sep. 2003.
[37] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using a generalized bilinear model," in *Proc. IEEE IGARSS*, Jul. 2011, pp. 1886–1889.
[38] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, "Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery," in *Proc. IEEE IGARSS*, Jul. 2011, pp. 3672–3675.
[39] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

Yoann Altmann (S'10–M'14) was born in Toulouse, France, in 1987. He received the Engineering degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, in 2010, and the Ph.D. degree from INP Toulouse in 2013. Since 2014, he has been with Heriot-Watt University, Edinburgh, as a Post-Doctoral Researcher. He conducts his research with the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences. His current research activities focus on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing and biomedical imaging.

Nicolas Dobigeon (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the Engineering degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, France, and the M.Sc. degree in signal processing and the Ph.D. and Habilitation Diriger des Recherches degrees in signal processing from the National Polytechnic Institute of Toulouse, in 2004, 2007, and 2012, respectively. From 2007 to 2008, he was a Post-Doctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been with the National Polytechnic Institute of Toulouse, INPENSEEIHT, University of Toulouse, where he is currently an Associate Professor. He conducts his research with the Signal and Communications Group, IRIT Laboratory, and he is also an Affiliated Faculty Member of the Telecommunications for Space and Aeronautics Cooperative Laboratory. His recent research has been focused on statistical signal and image processing, in particular, Bayesian inverse problems with applications to remote sensing, biomedical imaging, and genomics.

Steve McLaughlin (F'11) was born in Clydebank, Scotland, in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow and the Ph.D. degree from the University of Edinburgh in 1981 and 1990, respectively. From 1981 to 1984, he was a Development Engineer with the industry involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986, he was involved in the design and development of high-frequency data communication systems. In 1986, he joined the Department of Electronics and Electrical Engineering, University of Edinburgh, as a Research Fellow, where he studied the performance of linear adaptive algorithms in high-noise and nonstationary environments. In 1988, he joined the academic staff at Edinburgh, and from 1991 to 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he was awarded a Personal Chair in Electronic Communication Systems from the University of Edinburgh. In 2011, he joined Heriot-Watt University as a Professor of Signal Processing and the Head of the School of Engineering and Physical Sciences. His research interests include adaptive signal processing and nonlinear dynamical systems theory, and their applications to biomedical, energy, and communication systems. He is a fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology.

Jean-Yves Tourneret received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse and the Ph.D. degree from the National Polytechnic Institute of Toulouse, in 1989 and 1992, respectively. He is currently a Professor at ENSEEIHT, University of Toulouse, and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo methods. He has been involved in the organization of several conferences, including the 2002 European Conference on Signal Processing (Program Chair), the 2006 International Conference on Acoustics, Speech and Signal Processing (plenaries), the 2012 Statistical Signal Processing Workshop (international liaisons), the 2013 International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (local arrangement), and the 2014 Statistical Signal Processing Workshop (special sessions). He has been the Co-General Chair of the CIMI Workshop on Optimization and Statistics in Image Processing held in Toulouse in 2013, and a member of different technical committees, including the Signal Processing Theory and Methods Committee of the IEEE Signal Processing Society since 2010 and from 2001 to 2007. He was as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011 and has been serving as an Associate Editor for the *EURASIP Journal on Signal Processing* since 2013.